REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
FEBRUARY 2009	Conference Paper Preprint	October 2008 – February 2009

4. TITLE AND SUBTITLE
SPECTRAL MIXTURE MODELS FOR PARAMETER ESTIMATION (PREPRINT)

5a. CONTRACT NUMBER In House
5b. GRANT NUMBER N/A
5c. PROGRAM ELEMENT NUMBER 62702F

6. AUTHOR(S)
Alfredo Vega Irizarry

5d. PROJECT NUMBER 459E
5e. TASK NUMBER H9
5f. WORK UNIT NUMBER C2

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFRL/RIEC
525 Brooks Road
Rome NY 13441-4505

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFRL/RIEC
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
N/A

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TP-2009-31

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution unlimited PA# 88ABW-2009-0603 Date Cleared: 20-February-2009

13. SUPPLEMENTARY NOTES
This paper was rejected from the 2009 CISDA conference due to low relevance to conference topic. It was suggested to redo the introduction. This paper is reviewed by the author and will be resubmitted as a journal article in the IEEE Signal Processing. This is a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
A modification of the statistical technique known as Expectation Maximization can be applied for detecting and characterizing spectral activity. The novel technique adapts the Expectation Maximization to process histogram data using a signal model. The method provides signal clustering and parameter estimation.

15. SUBJECT TERMS
Spectral survey, Expectation Maximization, histogram data

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Alfredo Vega Irizarry
U	U	U	UU	5	19b. TELEPHONE NUMBER (*Include area code*) N/A

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Spectral Survey Using Spectral Mixture Models

Alfredo Vega Irizarry
Air Force Research Laboratory
525 Brooks Road
Rome, New York 13441–4505
Email: Alfredo.VegaIrizarry@rl.af.mil

***Abstract*—A modification of the statistical technique known as Expectation Maximization can be applied for detecting and characterizing spectral activity. The novel technique adapts the Expectation Maximization to process histogram data using a signal model. The method provides signal clustering and parameter estimation.**

I. Introduction

Conventional signal processing methods can be used to gain information of the spectral activity. Spectral survey methods can make use filtering, downconversion, frequency and phase locking techniques [5], simple statistics, and neural networks [5]. The methods can exploit unique features or use specialized processing techniques [4].

This paper is going to treat spectral survey as a pattern recognition problem. The proposed method provides clustering and parameter estimation capabilities. The development of the Spectral Mixture Models begins with a reinterpretation of the statistical parameters in terms of the telecommunications quantities. The similarities between the statistics and telecommunications are presented in Table I.

TABLE I
Comparison between Statistics and Telecommunications

Statistics	Telecommunications
Histogram	Spectrum
Clusters	Signals
Mean μ	Center Frequency f
Standard Deviation σ	Bandwidth b
Mixture probability α	Amplitude-bandwidth product α
Statistical sample	–
–	Signal Sample

The frequency spectrum provides multimodal histogram information. We wonder if a clustering algorithm such as the Gaussian Mixture Models can be use for gaining information about the activity in the spectrum. The idea sounds interesting, but it has a problem. There is no equivalent concept of a statistical sample in the telecommunications domain. Fortunately, it will be shown that it is possible to adapt the algorithm to process histogram data. The process will be discussed in this paper.

We will find out later that there is a second problem. The Gaussian model is not adequate for representing telecommunications signals. A more suitable model for digital modulation signals will be provided.

II. Expectation Maximization

The present discussion assumes that the reader is familiarized with the derivation of the Gaussian Mixture Models Algorithm which is an implementation of the Expectation Maximization using a Gaussian distribution.

The expectation of the log-likelihood $\mathcal{Q}(\vec{\theta})$ is defined in terms of the conditional expectation $E_{z'|\vec{y'}}$, the known data vector $\vec{y'}$, the missing data z' which identifies the cluster or mixture element, the parameter vector $\vec{\theta}$ and the joint probability of the data and the mixture element $p(\vec{y'}, z'; \vec{\theta})$.

$$\mathcal{Q}(\vec{\theta}) = E_{z'|\vec{y'}}\{p(\vec{y'}, z'; \vec{\theta})\} \tag{1}$$

The joint probability can be separated in two terms: the model $p(\vec{y'}|z'; \vec{\theta})$ and the mixture probability $P(z')$.

$$p(\vec{y'}, z'; \vec{\theta}) = p(\vec{y'}|z'; \vec{\theta})P(z'; \vec{\theta}) \tag{2}$$

The parameter vector $\vec{\theta}$ contains the following parameters: the mixture probability $\alpha = P(z'; \vec{\theta})$, the mean vector μ and the covariance matrix Σ. Maximizing the expectation with respect to these parameters provides optimal density functions that characterize the sample space.

$$\vec{\theta} = [\mu, \Sigma, \alpha] \tag{3}$$

In our spectral survey problem, the parameter vector consists of three quantities: the center frequency f, the bandwidth b and the amplitude-bandwidth product α.

$$\vec{\theta} = [f, b, \alpha] \tag{4}$$

The method under development can applied to multidimensional cases. An interesting application the application to the joint time-frequency domain. However, this discussion is out of the scope of our present paper. For simplicity, the problem is restricted to the one dimensional case which for the Gaussian Mixture Models consists at this moment of fictitious statistical samples.

$$\vec{y'} = y' \qquad \text{(scalar)} \tag{5}$$

III. Histogram Data

The Gaussian Mixture Models will be adapted to work with histogram data. The approach will consider two sets. The first set Y' describes J' statistical events that represent our known data. This is our input data $y' = y'_m$.

$$Y' = \{y'_m : j = 1, ..., M\} \tag{6}$$

A second set Y is defined as the histogram of the set Y'. Each partition j of the sample space is identified with a bin number y_j and the histogram count w_j. This set is what we would like our input to be $y = y_j$.

$$Y = \{(y_j, w_j) : j = 1, ..., J\} \tag{7}$$

We could assume that the elements of y'_n are sorted in ascending order. In such case, the elements of a bin y_j can be expressed by:

$$y_j = \{y'_n, y'_{n+1}, ..., y'_{n+w_j-1}\} \tag{8}$$

Associated with each bin y_j, there is also the missing data z_j which contains information about the mixture element or cluster index. (There is also a corresponding variable z'_m for each sample y'_m.) The cluster element is identified by an integer number from 1 to K, where K is the number of clusters.

$$1 \leq z_j \leq K \tag{9}$$

We will assume that every sample point within a histogram bin shares the same missing variable. In average, the statistical samples in a certain vicinity should belong to the same cluster. We can argue that if the bin size is small enough, this assumption will be true.

$$P(z'_n) = P(z'_{n+1}) = ... = P(z'_{n+w_j-1}) = \hat{P}(z_j = k) \tag{10}$$

An alternative and perhaps less confusing notation for the mixture probability is given by equation 11. This probability can be represented as a matrix with indexes for the bin and the cluster numbers.

$$\hat{P}(j, k) = \hat{P}(z_j = k) \tag{11}$$

The model must also be expressed in terms of the histogram variables. This step requires to define a new conditional probability for each bin $\hat{p}(y_j|z_j;\vec{\theta})$ as an average of the probabilities of all the statistical samples within a histogram bin.

$$\hat{p}(y_j|z_j = k;\vec{\theta}) = \frac{1}{w_j}\sum_{m=n}^{n+w_j-1} p(y'_m \subseteq y_j|z'_m = k;\vec{\theta}) \tag{12}$$

Now, we can proceed with the calculation of the joint probability density and the *a posteriori* probability needed by the Expectation Maximization. The term w_j is canceled resulting in an expression that is similar to the original Bayesian estimate.

$$\hat{p}(y_j, z_j = k;\vec{\theta}) = \frac{1}{w_j}\sum_{m=n}^{n+w_j-1} p(y'_m \subseteq y_j|z'_m = k)\hat{P}(z_j = k) \tag{13}$$

$$\hat{p}(y_j;\vec{\theta}) = \sum_{k=1}^{K} \hat{p}(y_j, z_j = k;\vec{\theta})\hat{P}(z_j = k) \tag{14}$$

$$\hat{p}(z_j = k|y_j;\vec{\theta}) = \frac{\hat{p}(y_j, z_j = k;\vec{\theta})}{\hat{p}(y_j;\vec{\theta})} \tag{15}$$

The histogram data will be incorporated into the conditional expectation of the log-likelihood. Each probability of the histogram bin data is assumed to be a independent event.

The variable w_j indicates the number of times that the joint probability shown in equation 13 appears in the likelihood expression. This is equivalent to raise each probability term to the w_j power. The term $\hat{P}(z_j;\vec{\theta})^w$ will be treated as a single term for convenience.

$$\alpha = P(z_j;\vec{\theta}) = \hat{P}(z_j;\vec{\theta})^w \tag{16}$$

The variable w_j was originally defined as an integer, but by looking at equation 17, we notice that the range of w can be extended to any a real positive number. This quantity will be used to represent the amplitude of the spectrum.

$$\mathcal{Q}(\vec{\theta}) = E_{z|\vec{y}}\{\hat{p}(\vec{y}|z;\vec{\theta})^w P(z;\vec{\theta})\} \tag{17}$$

The maximization process requires the calculation of the derivatives of with respect to each parameter: f, b and α and finding the roots. The roots are the new updated parameters. Finding the roots may require implementing methods such as the Newton's algorithm.

$$\nabla_\theta \mathcal{Q}(\vec{\theta}) = 0 \tag{18}$$

The parameter updates (M-step) are used to calculate the *a priori* probabilities (E-step) using the Bayesian formulas that have been derived. The process is repeated until the process converges to some local maximum.

$$\vec{\theta} \approx \vec{\theta}_{max} \tag{19}$$

The parameter vector was relabeled to use the spectral survey parameters. After running several trials, the new algorithm was unable to converge to meaningful parameters. The algorithm processed a mixture of MPSK and MFSK signals. The equations and algorithm appeared to be implemented properly. Using a Gaussian model was not an good choice. Adjacent clusters interfere with each other due to the wide transition bandwidth. The model also lacks of a flat passband.

IV. Clipped Gaussian Model

Our telecommunication signal requires a bandwidth efficient model. The model must incorporate a profile that approximates an ideal filter response in the frequency domain. Our first choice would be to use a Butterworth filter model. Unfortunately, the use of this model results in complex analytical equations during the maximization process. Our desire is to find bandwidth efficient models that are computationally efficient.

A distribution known as Clipped Gaussian will be used for our model [1]. The model provides a flat pass band and a steep transition bandwidth. The model looks appealing because the log-likelihood simplifies the product of the exponential function. The implementation of the algorithm using this model produced promising results.

$$\hat{p}(y|z;\vec{\theta}) = \frac{N}{b \cdot \Gamma(\frac{1}{N})}\ exp\left(\left(\frac{y-f}{b/2}\right)^N\right) \tag{20}$$

Where, N is an even positive integer and $\Gamma(x)$ is the incomplete Gamma function. Selecting $b = 2\sqrt{2}\sigma$ and $N = 2$ results in the Gaussian distribution, but this is not a good choice. Our signal model was built using $N = 8$.

V. Overestimation

The number signals in the spectrum is assumed to be unknown. We hope that initializing the algorithm with many mixtures will result in overestimation. This will allow us to capture more details in the frequency spectrum.

The overestimation in the Spectral Mixture Models generates suboptimal solutions which will be referred as subbands. The final solutions will be referred as bands or signals. The algorithm will require an additional step for recombining adjacent subbands that belong to the same band. The recombination can make use of the parameters to determine the proximity with two adjacent subbands. Recombination should be done when the algorithm has converged to some suboptimal solution.

VI. Controlling Convergence

The Clipped Gaussian model has a singularity at $b = 0$. To avoid convergence problems, we can execute a rule immediately after updating the parameters. The rule will limit the growth of b. The rule can be implemented with a simple if statement.

$$if(b < b_{min})\ then\ b_j = b_{min}\ endif$$

Other useful way to control the convergence of the algorithm is by reserving one mixture element for noise floor characterization. The bandwidth of one mixture is forced to be wide enough so it is force to converge to the receiver noise floor.

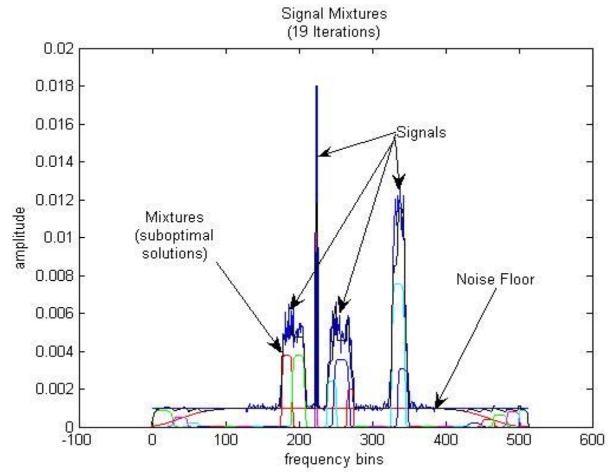

Fig. 1. Convergence of 40 spectral mixtures after 19 iterations. Signals from left to right: QPSK, tone, QPSK, FSK-2 and noise floor.

VII. Simulations

Tables II to VII contain the results of several Monte Carlo simulations. The simulations were conducted using combination of center frequencies: 2, 4 and 6 kHz; bandwidth: 1 kHz; and various signal-to-noise ratios (SNR): 40, 30, 20, 10, 5 and 2 dB. The algorithm was initialized with uniformly spaced mixtures. The initial bandwidths were uniform except for the bandwidth of one mixture that is forced to converge to the noise floor of the receiver. The algorithm was stopped after 25 iterations. The signal under test is a QPSK type sampled at 16 kHz. The message is a random sequence that contains between 400 and 500 symbols. The true number of cycle is a multiple of the total number of channels in the analysis filter used to estimate the frequency spectrum. A total of 256 filter channels were used in all the simulations.

VIII. Conclusion

Expectation Maximization can be use for spectral survey by adapting the algorithm in two ways. First, the algorithm needs to process histogram-like data such as the frequency spectrum or spectrogram. Second, the model needs to be representative of a signal. A display of the spectral mixtures is shown in Figure 1.

The concept can be refined to produce better estimates. Better combination rules can be implemented. Initial conditions can be modified to acelerate the convergence. It is important to notice that the parameter b was not meant to be a 3dB bandwidth estimate. A better bandwidth estimates could be produced by mapping b to the real 3dB bandwidth.

Potential applications of this method include cognitive radios, broadband receivers and measurement equipment.

Acknowledgment

The author would like to thank Dr. Andrew Noga of the Air Force Research Laboratory for his invaluable guidance and assistance in reviewing the theory presented in this paper.

TABLE II
CENTER FREQUENCY ESTIMATES FOR VARIOUS SNR LEVELS. TRUE PARAMETERS CENTER FREQUENCY = 2 KHZ, BANDWIDTH = 1 KHZ.

$Fc = 2kHz$	Center Frequency Estimate (Hz)		
SNR	Mean	Bias	Std. Dev.
2	2005.48	-5.48	437.81
5	2057.59	-57.59	339.73
10	2010.82	-10.82	278.32
20	2015.63	-15.63	166.07
30	2024.53	-24.53	138.25
40	2022.31	-22.31	135.74

TABLE III
CENTER FREQUENCY ESTIMATES FOR VARIOUS SNR LEVELS. TRUE PARAMETERS CENTER FREQUENCY = 4 KHZ, BANDWIDTH = 1 KHZ.

$Fc = 4kHz$	Center Frequency Estimate (Hz)		
SNR	Mean	Bias	Std. Dev.
2	4303.15	-303.15	70.04
5	4026.37	-26.37	144.73
10	4020.42	-15.80	52.56
20	4015.80	-15.65	8.15
30	4015.65	-15.70	4.80
40	4015.70	-15.70	2.91

TABLE IV
CENTER FREQUENCY ESTIMATES FOR VARIOUS SNR LEVELS. TRUE PARAMETERS CENTER FREQUENCY = 6 KHZ, BANDWIDTH = 1 KHZ.

$Fc = 6kHz$	Center Frequency Estimate (Hz)		
SNR	Mean	Bias	Std. Dev.
2	5339.43	660.56	144.27
5	5957.93	42.06	42.23
10	6011.36	-14.84	5.40
20	6014.84	-15.46	10.32
30	6015.46	-15.68	4.74
40	6015.68	-15.68	3.55

TABLE V
BANDWIDTH ESTIMATES FOR VARIOUS SNR LEVELS. TRUE PARAMETERS CENTER FREQUENCY = 2 KHZ, BANDWIDTH = 1 KHZ.

$Fc = 2kHz$	Bandwidth Estimate (Hz)		
SNR	Mean	Bias	Std. Dev.
2	1063.18	-63.18	643.80
5	1249.49	-249.49	603.34
10	1409.90	-409.90	472.68
20	1670.68	-670.68	303.80
30	1747.25	-747.25	255.34
40	1796.64	-760.98	250.65

TABLE VI
BANDWIDTH ESTIMATES FOR VARIOUS SNR LEVELS. TRUE PARAMETERS CENTER FREQUENCY = 4 KHZ, BANDWIDTH = 1 KHZ.

$Fc = 4kHz$	Bandwidth Estimate (Hz)		
SNR	Mean	Bias	Std. Dev.
2	1255.90	-255.90	822.41
5	1332.06	-332.06	677.71
10	1454.18	-454.18	450.95
20	1652	-652	273.95
30	1709.18	-709.18	233.25
40	1738.41	-738.41	211.03

TABLE VII
BANDWIDTH ESTIMATES FOR VARIOUS SNR LEVELS. TRUE PARAMETERS CENTER FREQUENCY = 6 KHZ, BANDWIDTH = 1 KHZ.

$Fc = 6kHz$	Bandwidth Estimate (Hz)		
SNR	Mean	Bias	Std. Dev.
2	1100.72	-100.72	671.86
5	1167.12	-167.12	593.57
10	1380.17	-380.17	488.73
20	1649.5	-649.5	319.96
30	1752.25	-752.25	248.30
40	1764.62	-764.62	239.42

REFERENCES

[1] US Patent 6,269,334 (Basu et al.)
[2] US Patent 4,827,225 (Lee)
[3] US Patent 7,092,436 (Ma et al.)
[4] US Patent 5,257,211 (Noga)
[5] US Patent 4,904,930 (Nicholas)
[6] US Patent 6,819,731 (Nooralahiyan)
[7] US Patent 6,002,721 (Said)
[8] R. Duda, *Pattern Classification*, 2nd Edition, John Wiley and Sons, 2001
[9] T. Moon, *The Expectation Maximization Algorithm*, IEEE Signal Processing Magazine; Vol 13, No. 6, November 1996, pg. 47-64.
[10] A. Noga, *Methods for Discrete-Time Phase and Frequency Synchronization*; DTIC, AFRL In-House Report AFRL-IF-RS-TR-1999-133, May 1995
[11] S. Theodoridis and K. Koutroumbas, *Pattern Recognition* 3rd Edition, Elsevier, 2006